

■ part 1

Introduction to Databases ■

chapter 1 Databases and Database Users 3

1.1 Introduction 4

1.2 An Example 6 1.3 Characteristics of the Database Approach 10 1.4 Actors on the Scene 15 1.5 Workers behind the Scene 17 1.6 Advantages of Using the DBMS Approach 17 1.7 A Brief History of Database Applications 23 1.8 When Not to Use a DBMS 27 1.9 Summary 27 Review Questions 28 Exercises 28 Selected Bibliography 29

chapter 2 Database System Concepts and Architecture 31

2.1 Data Models, Schemas, and Instances 32

2.2 Three-Schema Architecture and Data Independence 36 2.3 Database Languages and Interfaces 38

2.4 The Database System Environment 42 2.5 Centralized and Client/Server Architectures for DBMSs 46

2.6 Classification of Database Management Systems 51 2.7 Summary 54 Review Questions 55 Exercises 55 Selected Bibliography 56

xviii Contents

■ part 2

Conceptual Data Modeling and Database Design ■

chapter 3 Data Modeling Using the Entity–Relationship (ER) Model 59

3.1 Using High-Level Conceptual Data Models for Database Design 60

3.2 A Sample Database Application 62 3.3 Entity Types, Entity Sets, Attributes, and Keys 63 3.4 Relationship Types, Relationship Sets, Roles, and Structural Constraints 72

3.5 Weak Entity Types 79 3.6 Refining the ER Design for the COMPANY Database 80 3.7 ER Diagrams, Naming Conventions, and Design Issues 81 3.8 Example of Other Notation: UML Class Diagrams 85 3.9 Relationship Types of Degree Higher than Two 88 3.10 Another Example: A UNIVERSITY Database 92 3.11 Summary 94 Review Questions 96 Exercises 96 Laboratory Exercises 103 Selected Bibliography 104

chapter 4 The Enhanced Entity–Relationship (EER) Model 107

4.1 Subclasses, Superclasses, and Inheritance 108

4.2 Specialization and Generalization 110 4.3 Constraints and Characteristics of Specialization and Generalization 113

4.4 Modeling of UNION Types Using Categories 120 4.5 A Sample UNIVERSITY EER Schema, Design Choices, and Formal Definitions 122

4.6 Example of Other Notation: Representing Specialization and Generalization in UML Class Diagrams 127

4.7 Data Abstraction, Knowledge Representation, and Ontology Concepts 128

4.8 Summary 135 Review Questions 135 Exercises 136 Laboratory Exercises 143 Selected Bibliography 146 Contents xix

■ part 3

The Relational Data Model and SQL ■

chapter 5 The Relational Data Model and Relational Database Constraints 149

5.1 Relational Model Concepts 150

5.2 Relational Model Constraints and Relational Database Schemas 157 5.3 Update Operations, Transactions, and Dealing with Constraint Violations 165

5.4 Summary 169 Review Questions 170 Exercises 170 Selected Bibliography 175

chapter 6 Basic SQL 177

6.1 SQL Data Definition and Data Types	179	6.2 Specifying Constraints in SQL	184	6.3 Basic Retrieval Queries in SQL	187
6.4 INSERT, DELETE, and UPDATE Statements in SQL	198	6.5 Additional Features of SQL	201	6.6 Summary	202
Review Questions	203	Exercises	203	Selected Bibliography	205
chapter 7 More SQL: Complex Queries, Triggers, Views, and Schema Modification	207	7.1 More Complex SQL Retrieval Queries	207	7.2 Specifying Constraints as Assertions and Actions as Triggers	225
7.3 Views (Virtual Tables) in SQL	227	7.4 Schema Change Statements in SQL	232	7.5 Summary	234
Review Questions	236	Exercises	236	Selected Bibliography	238
chapter 8 The Relational Algebra and Relational Calculus	239	8.1 Unary Relational Operations: SELECT and PROJECT	241	8.2 Relational Algebra Operations from Set Theory	246
8.3 Binary Relational Operations: JOIN and DIVISION	251	8.4 Additional Relational Operations	259	8.5 Examples of Queries in Relational Algebra	265
8.6 The Tuple Relational Calculus	268	8.7 The Domain Relational Calculus	277	8.8 Summary	279
Review Questions	280	Exercises	281	Laboratory Exercises	286
Selected Bibliography	288	chapter 9 Relational Database Design by ER- and EER-to-Relational Mapping	289	9.1 Relational Database Design Using ER-to-Relational Mapping	290
9.2 Mapping EER Model Constructs to Relations	298	9.3 Summary	303	Review Questions	303
Exercises	303	Laboratory Exercises	305	Selected Bibliography	306
■ part 4		Database Programming Techniques ■		chapter 10 Introduction to SQL Programming Techniques	309
10.1 Overview of Database Programming Techniques and Issues	310	10.2 Embedded SQL, Dynamic SQL, and SQLJ	314	10.3 Database Programming with Function Calls and Class Libraries: SQL/CLI and JDBC	326
10.4 Database Stored Procedures and SQL/PSM	335	10.5 Comparing the Three Approaches	338	10.6 Summary	339
Review Questions	340	Exercises	340	Selected Bibliography	341
chapter 11 Web Database Programming Using PHP	343	11.1 A Simple PHP Example	344	11.2 Overview of Basic Features of PHP	346
xx Contents	11.3	Overview of PHP Database Programming	353	11.4 Brief Overview of Java Technologies for Database Web Programming	358
11.5 Summary	358	Review Questions	359	Exercises	359
Selected Bibliography	359	■ part 5		Object, Object-Relational, and XML: Concepts, Models, Languages, and Standards ■	
chapter 12 Object and Object-Relational Databases	363	12.1 Overview of Object Database Concepts	365	12.2 Object Database Extensions to SQL	379
12.3 The ODMG Object Model and the Object Definition Language ODL	386	12.4 Object Database Conceptual Design	405	12.5 The Object Query Language OQL	408
12.6 Overview of the C++ Language Binding in the ODMG Standard	417	12.7 Summary	418	Review Questions	420
Exercises	421	Selected Bibliography	422	chapter 13 XML: Extensible Markup Language	425
13.1 Structured, Semistructured, and Unstructured Data	426	13.2 XML Hierarchical (Tree) Data Model	430	13.3 XML Documents, DTD, and XML Schema	433
13.4 Storing and Extracting XML Documents					

from Databases 442

13.5 XML Languages 44313.6 Extracting XML Documents from Relational Databases 44713.7 XML/SQL: SQL Functions for Creating XML Data 45313.8 Summary 455Review Questions 456Exercises

456Selected Bibliography 456 Contents xxi■ part 6

Database Design Theory and Normalization ■

chapter 14 Basics of Functional Dependencies and Normalization for Relational Databases 459

14.1 Informal Design Guidelines for Relation

Schemas 461

14.2 Functional Dependencies 47114.3 Normal Forms Based on Primary Keys 47414.4 General Definitions of Second and Third Normal

Forms 483

14.5 Boyce-Codd Normal Form 48714.6 Multivalued Dependency and Fourth

Normal Form 491

14.7 Join Dependencies and Fifth Normal Form 49414.8 Summary 495Review Questions 496Exercises

497Laboratory Exercises 501Selected Bibliography 502

chapter 15 Relational Database Design Algorithms

and Further Dependencies 503

15.1 Further Topics in Functional Dependencies: Inference Rules,

Equivalence, and Minimal Cover 505

15.2 Properties of Relational Decompositions 51315.3 Algorithms for Relational Database Schema

Design 519

15.4 About Nulls, Dangling Tuples, and Alternative Relational

Designs 523

15.5 Further Discussion of Multivalued Dependencies

and 4NF 527

15.6 Other Dependencies and Normal Forms 53015.7 Summary 533Review Questions 534Exercise

s 535Laboratory Exercises 536Selected Bibliography 537xxii Contents■ part 7

File Structures, Hashing, Indexing, and Physical

Database Design ■

chapter 16 Disk Storage, Basic File Structures,

Hashing, and Modern Storage Architectures 541

16.1 Introduction 542

16.2 Secondary Storage Devices 54716.3 Buffering of Blocks 55616.4 Placing File Records on Disk

56016.5 Operations on Files 56416.6 Files of Unordered Records (Heap Files) 56716.7 Files of Order

ed Records (Sorted Files) 56816.8 Hashing Techniques 57216.9 Other Primary File Organizations 5

8216.10 Parallelizing Disk Access Using RAID

Technology 584

16.11 Modern Storage Architectures 58816.12 Summary 592Review Questions 593Exercises 595S

electd Bibliography 598

chapter 17 Indexing Structures for Files and Physical

Database Design 601

17.1 Types of Single-Level Ordered Indexes 602

17.2 Multilevel Indexes 61317.3 Dynamic Multilevel Indexes Using B-Trees

and B

+Trees 617

17.4 Indexes on Multiple Keys 63117.5 Other Types of Indexes 63317.6 Some General Issues Concer

ning Indexing 63817.7 Physical Database Design in Relational

Databases 643

17.8 Summary 646Review Questions 647Exercises 648Selected Bibliography 650 Contents xxiii■ p

art 8

Query Processing and Optimization ■

chapter 18 Strategies for Query Processing 655

18.1 Translating SQL Queries into Relational Algebra

and Other Operators 657

18.2 Algorithms for External Sorting	660
18.3 Algorithms for SELECT Operation	663
18.4 Implementing the JOIN Operation	668
18.5 Algorithms for PROJECT and Set Operations	676
18.6 Implementing Aggregate Operations and Different Types of JOINS	678
18.7 Combining Operations Using Pipelining	681
18.8 Parallel Algorithms for Query Processing	683
18.9 Summary	688
Review Questions	688
Exercises	689
Selected Bibliography	689
chapter 19 Query Optimization	691
19.1 Query Trees and Heuristics for Query Optimization	692
19.2 Choice of Query Execution Plans	701
19.3 Use of Selectivities in Cost-Based Optimization	710
19.4 Cost Functions for SELECT Operation	714
19.5 Cost Functions for the JOIN Operation	717
19.6 Example to Illustrate Cost-Based Query Optimization	726
19.7 Additional Issues Related to Query Optimization	728
19.8 An Example of Query Optimization in Data Warehouses	731
19.9 Overview of Query Optimization in Oracle	733
19.10 Semantic Query Optimization	737
19.11 Summary	738
Review Questions	739
Exercises	740
Selected Bibliography	740
xxiv Contents ■ part 9	
Transaction Processing, Concurrency Control, and Recovery ■	
chapter 20 Introduction to Transaction Processing Concepts and Theory	745
20.1 Introduction to Transaction Processing	746
20.2 Transaction and System Concepts	753
20.3 Desirable Properties of Transactions	757
20.4 Characterizing Schedules Based on Recoverability	759
20.5 Characterizing Schedules Based on Serializability	763
20.6 Transaction Support in SQL	773
20.7 Summary	776
Review Questions	777
Exercises	777
Selected Bibliography	779
chapter 21 Concurrency Control Techniques	781
21.1 Two-Phase Locking Techniques for Concurrency Control	782
21.2 Concurrency Control Based on Timestamp Ordering	792
21.3 Multiversion Concurrency Control Techniques	795
21.4 Validation (Optimistic) Techniques and Snapshot Isolation	798
21.5 Granularity of Data Items and Multiple Granularity Locking	800
21.6 Using Locks for Concurrency Control in Indexes	805
21.7 Other Concurrency Control Issues	806
21.8 Summary	807
Review Questions	808
Exercises	809
Selected Bibliography	810
chapter 22 Database Recovery Techniques	813
22.1 Recovery Concepts	814
22.2 NO-UNDO/REDO Recovery Based on Deferred Update	821
22.3 Recovery Techniques Based on Immediate Update	823
22.4 Shadow Paging	826
22.5 The ARIES Recovery Algorithm	827
22.6 Recovery in Multidatabase Systems	831
22.7 Database Backup and Recovery from Catastrophic Failures	832
22.8 Summary	833
Review Questions	834
Exercises	835
Selected Bibliography	838
■ part 10	
Distributed Databases, NOSQL Systems, and Big Data ■	
chapter 23 Distributed Database Concepts	841
23.1 Distributed Database Concepts	842
23.2 Data Fragmentation, Replication, and Allocation Techniques for Distributed Database Design	847
23.3 Overview of Concurrency Control and Recovery in Distributed Databases	854

23.4 Overview of Transaction Management in Distributed Databases	857
23.5 Query Processing and Optimization in Distributed Databases	859
23.6 Types of Distributed Database Systems	865
23.7 Distributed Database Architectures	868
23.8 Distributed Catalog Management	875
23.9 Summary	876
Review Questions	877
Exercises	878
Selected Bibliography	880
chapter 24 NOSQL Databases and Big Data Storage Systems	883
24.1 Introduction to NOSQL Systems	884
24.2 The CAP Theorem	888
24.3 Document-Based NOSQL Systems and MongoDB	890
24.4 NOSQL Key-Value Stores	895
24.5 Column-Based or Wide Column NOSQL Systems	900
24.6 NOSQL Graph Databases and Neo4j	903
24.7 Summary	909
Review Questions	909
Selected Bibliography	910
Contents	xxvi
chapter 25 Big Data Technologies Based on MapReduce and Hadoop	911
25.1 What Is Big Data?	914
25.2 Introduction to MapReduce and Hadoop	916
25.3 Hadoop Distributed File System (HDFS)	921
25.4 MapReduce: Additional Details	926
25.5 Hadoop v2 alias YARN	936
25.6 General Discussion	944
25.7 Summary	953
Review Questions	954
Selected Bibliography	956
■ part 11	
Advanced Database Models, Systems, and Applications	■
chapter 26 Enhanced Data Models: Introduction to Active, Temporal, Spatial, Multimedia, and Deductive Databases	961
26.1 Active Database Concepts and Triggers	963
26.2 Temporal Database Concepts	974
26.3 Spatial Database Concepts	987
26.4 Multimedia Database Concepts	994
26.5 Introduction to Deductive Databases	999
26.6 Summary	1012
Review Questions	1014
Exercises	1015
Selected Bibliography	1018
chapter 27 Introduction to Information Retrieval and Web Search	1021
27.1 Information Retrieval (IR) Concepts	1022
27.2 Retrieval Models	1029
27.3 Types of Queries in IR Systems	1035
27.4 Text Preprocessing	1037
27.5 Inverted Indexing	1040
27.6 Evaluation Measures of Search Relevance	1044
27.7 Web Search and Analysis	1047
Contents	xxvii
27.8 Trends in Information Retrieval	1057
27.9 Summary	1063
Review Questions	1064
Selected Bibliography	1066
chapter 28 Data Mining Concepts	1069
28.1 Overview of Data Mining Technology	1070
28.2 Association Rules	1073
28.3 Classification	1085
28.4 Clustering	1088
28.5 Approaches to Other Data Mining Problems	1091
28.6 Applications of Data Mining	1094
28.7 Commercial Data Mining Tools	1094
28.8 Summary	1097
Review Questions	1097
Exercises	1098
Selected Bibliography	1099
chapter 29 Overview of Data Warehousing and OLAP	1101
29.1 Introduction, Definitions, and Terminology	1102
29.2 Characteristics of Data Warehouses	1103
29.3 Data Modeling for Data Warehouses	1105
29.4 Building a Data Warehouse	1111
29.5 Typical Functionality of a Data Warehouse	1114
29.6 Data Warehouse versus Views	1115
29.7 Difficulties of Implementing Data Warehouses	1116
29.8 Summary	1117
Review Questions	1117
Selected Bibliography	1118
■ part 12	
Additional Database Topics: Security	■
chapter 30 Database Security	1121
30.1 Introduction to Database Security Issues	1122
30.2 Discretionary Access Control Based on Granting and Revoking Privileges	1129
30.3 Mandatory Access Control and Role-Based Access Control for Multilevel Security	1134
Contents	xxviii
30.4 SQL Injection	1143
30.5 Introduction to Statistical Database Security	1146
30.6 Introduction to Flow Control	1147
30.7 Encryption and Public Key Infrastructures	1149
30.8 Privacy Issues and Preservation	1153
30.9 Challenges to Maintaining Database Security	1154
30.10 Oracle Label-Based Security	